Tax Cheating in Synchronic Online Experiments: A Pre-Registration Plan *

Raymond M. Duch
raymond.duch@nuffield.ox.ac.uk
Nuffield College
University of Oxford

Akitaka Matsuo
akitaka.matsuo@nuffield.ox.ac.uk
Nuffield College
University of Oxford

March 23, 2016

*Centre for Experimental Social Sciences, Nuffield College, University of Oxford. March, 2016.

Abstract

Recent lab experimental evidence reported by Duch and Solaz (2016) suggest that: individuals who demonstrate high levels of ability or effort, who typically are rich, are much less likely than others to comply with taxation. A causal mechanism contributing to tax cheating appears to be ability in addition to wealth per se.

There is overwhelming evidence that subjects who perform better on the real effort tasks cheat more – simply performing better, in our experiments, results in greedier behaviour. Experimental treatments were implemented in order to explore alternative causal mechanism: As the cost of compliance rises (a function of tax rate and earnings) they find an increase in cheating. But the higher levels of cheating by able versus less able types persists in high and low tax regimes. They find no experimental evidence that when earnings are associated with luck or status that this correlation between performance or ability and cheating moderates; no evidence that more redistributive treatments affected intrinsic motivations of those with high ability.

Duch and Solaz (2016) find that intrinsic motivation for complying with taxation is very asymmetric. Those who exhibit high ability, and hence are more likely to be rich, realise much less intrinsic benefits from complying with taxation than is the case for those with lower ability who are more likely to be poor.

We are pre-registering an experiment that is designed to replicate exactly the lab experiment of Duch and Solaz (2016) in an online synchronic experiment with 500 M-Turk subjects.

1 Online Tax Compliance Experiments

We will implement one of the treatments Duch and Solaz (2016) implemented in their lab experiments – the baseline treatment. Subjects are paid at the end of the online experiment, and do not receive feedback about earnings until the end of the experiment. Participants receive visual instructions on their screens at the beginning of each module (unlike the lab experiments instructions are NOT read and explained aloud).

The tax treatment consists of ten rounds. Table 1 summarises the treatments that are implemented in the main module of the experiment. Prior to the tax treatments, participants are randomly assigned to groups of four and we follow a partner matching. Thus, the composition of each group remains unchanged for the tax treatment module. Each round of the module is divided into two stages. In the first stage subjects perform a real effort task. This task consist of computing a series of additions in one minute. Their Preliminary Gains depend on how many correct answers they provide, getting 150 ECUs for each correct answer.

Once subjects have received information concerning their Preliminary Gains, participants are asked to declare these gains. A certain percentage or "tax" (that depends on the treatment) of these Declared Gains is then deducted from their Preliminary Gains.[1] These deductions are then evenly divided amongst the members of the group. Note that in each session the tax rate is consistent. The tax treatments implemented on the online experiments are the following: 10% and 30%.

In the equal salary (*Baseline*) treatment Duch and Solaz (2016) paid the lab subjects the same payment for correct answers to the real effort test (10 pence) – in our online version subjects receive 15 cents (USD). Salaries are strictly tied to performance.

In each module there is a certain probability that the Declared Gains are compared with the actual Preliminary Gains in order to verify these two amounts correspond. In

[1] We explicitly avoid framing the game in terms of "taxes". Subjects are told that a deduction (rather than a "tax") would be applied to earnings.

Table 1: Summary of Online Tax Compliance Experimental Treatments

Session	Participants	Groups	Tax Rate	AR Block 1	AR Block 2	Treatment
1	100	6	10%	0%	10%	Baseline
2	100	6	30%	0%	10%	Baseline

one module the probability is 0%, while this probability changes to 10% in our future sessions.

At the end of each round participants are informed of their Preliminary and Declared gains; the amount they receive from the deductions in their group; and the earnings in the round. At the end of each tax session one of the ten rounds is chosen at random, and their earnings are based on their profit for that round.

At the end of the experiment their earnings in ECUs are converted to USD at the exchange rate 1000ECUs = 1$ (this compares to 300ECUs = 1£in the Duch and Solaz (2016) lab experiment). Participants are then asked to answer a questionnaire, which consists on an Integrity Test, and a series of socio-demographic questions.

The experiment is programmed in NodeGame (Balietti 2015). A copy of the instructions along with screen shots can be found in the Appendix.

2 Conjectures Regarding Cheating

The experimental treatments are designed to identify the mechanisms that explain cheating – they correspond to conjectures regarding tax compliance outlined initially in Duch and Solaz (2016).

Ability. Our central argument is that high ability types receive significantly lower intrinsic payoffs from tax compliance than do low ability types. We argue consistent with

Duch and Solaz (2016) that in general those with high levels of ability or skill are more likely to engage in unethical behaviour – cheating at one's taxes is an illustration of such unethical behaviour.

Price of intrinsic payoffs. The intrinsic payoffs individuals might realise from tax compliance obviously come at an "extrinsic" cost. Individuals in the population, regardless of whether they are of high or low ability, share a similar, presumably downward sloping, demand function for "intrinsic" rewards.

Clearly, the costs of the intrinsic payoffs from tax compliance will be highly correlated with one's ability. High ability types typically face high costs associated with indulging their prosocial instincts. The two tax rate treatments described earlier are designed to help tease out the relative importance of these two determinants of tax compliance: ability versus costs. We observe low and high ability types at different prevailing tax rates; and for any compliance cost level, we observe how ability levels affect rates of cheating.

Winners versus Losers. We argue that individuals with high ability are more likely to cheat because of their type. High performance types are typically "winners" but as we pointed out earlier its not the outcome of any particular competition per se that causes these individuals to cheat. An alternative mechanism is that cheating behaviour is simply triggered by these discrete events that identify a winner or loser. Our experimental design goes to considerable length to ensure we are not confounding "winning" with high performance type.

The argument about "winning" and cheating distinctly invokes knowledge of how individuals perform relative to others in the "competition", or of how one's earnings compare to other's earnings. In our experiments, we control access to this information which helps us isolate the effect of performance type as opposed to the effect of one's "winner" versus "loser" status in a particular competition.

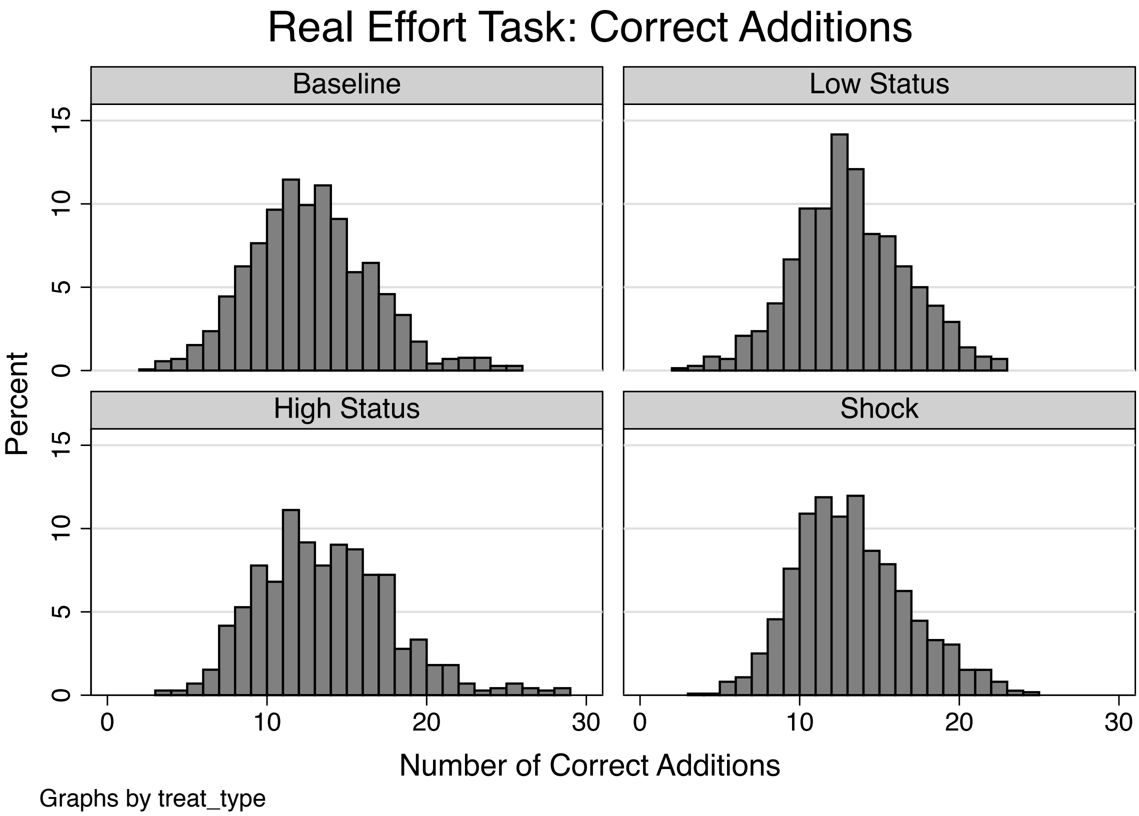
Secondly, in our experiments we observe cheating behaviour over multiple rounds of the same tax compliance game. We can think of each round as a competition. And we can observe the extent to which our high and low performance types exhibit consistent cheating behaviour. While it is true that we expect high performance types to do well on average, we will observe rounds in which they perform below average. If winning and losing in particular competitions affects cheating behaviour (as opposed to being affected by ones general type) then we should observe significant variations subject cheating behaviour when this occurs.

3 Analysis Plan for Online Synchronic Tax Compliance Experiment

3.1 Effort

Income from the real effort tasks (additions) is entirely determined by performance – each subject will receive 15 cents (150 ECUs) per correct addition. Figure 1 summarises the performance of the subjects in the original lab experiment. The mean correct additions is very similar in the across the different lab experiment treatments: 12.2 in the baseline; 12.7 in the low status; 13.2 in the high status; and 12.8 in the shock treatment. There was also no indication that effort was conditioned on the tax rates – the average correct additions for the 10%, 20%, and 30% tax rates, respectively, were 12.7, 11.8, and 12.1. We will exactly replicate this Figure with the performance of our online subjects (along with the comparative lab experiments results)

Figure 1: Real Effort Tasks: Correct Additions in Baseline, Status and Shock Treatments

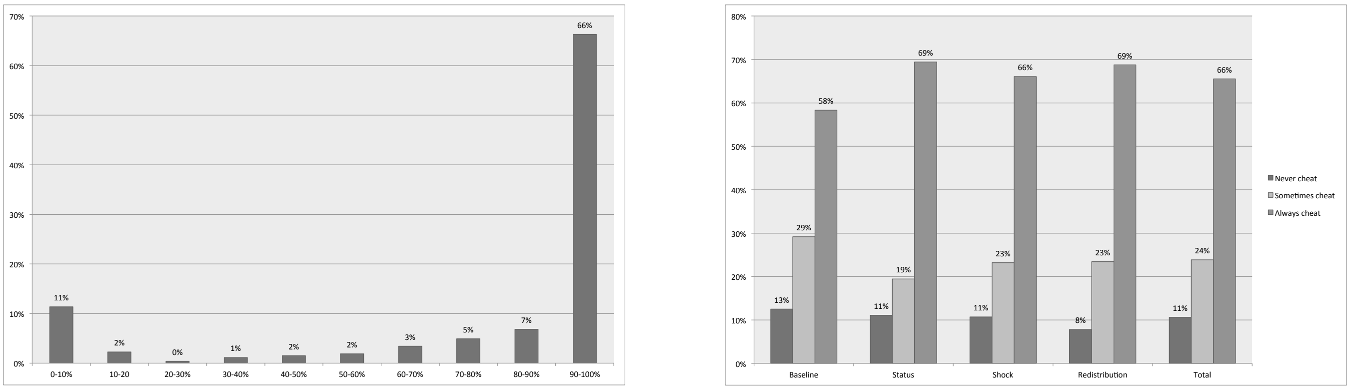


3.2 Cheating

The outcome variable of interest in this experiment will be tax compliance. In order to ensure subjects had incentives to cheat we adopted an audit rate of zero and 10% hence subjects were either not penalised at all for cheating or the probability is relatively low. Revenue collected from these taxes will be distributed equally amongst subjects and hence there are no social gains (or losses) associated with compliance. As a result the equilibrium choice for all subjects is to report zero earnings.

Figure 2 summarises the subjects' cheating behaviour from the Duch and Solaz (2016) lab experiment. The left graph in Figure 2 presents the frequency of subjects' average ratio of non-declared to total earnings. About two-thirds of the subjects are cheating virtually in every round. And about 10 percent never cheat. This Figure will be replicated for the analysis of the online synchronic experiment – again we'll compare the lab with the online results.

Figure 2: Subjects' Non-declared to Total Earnings

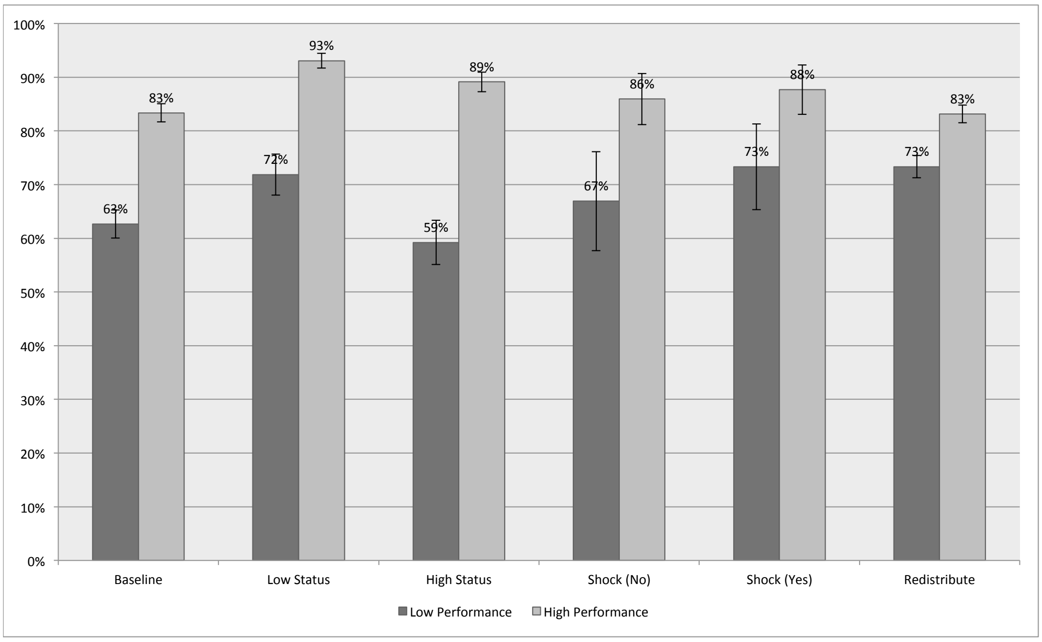


(a) Frequency Average Non-declared to Total Earnings

(b) Frequency: Alway, Sometimes, Never Cheat

Figure 3 presents the lab experiment results for low and high performance types. A similar Figure (with comparative results from the lab experiment) will be generated for the online synchronic experiment. We will define low and high performance types according to whether their average performance was above or below the median performance on the real effort tasks under each treatment (which in the lab experiment was either 11 or 12 correct additions depending on the treatment). There is no question here that performance type matters – high performance types cheat. And the results are quite robust to treatment: High performance types cheat more regardless of treatment. In the Baseline treatment, there is a difference of about 20 percent in cheating – high performance types do not declare about 85 percent of their income while low performance types cheat at a rate of about 65 percent. Note that only the Baseline treatment will be replicated in the online synchronic experiment. Our expectation is that the same correlation between performance and cheating will hold with the online experiment.

Figure 3: Average Ratio of Non-declared Earnings over Performance Type and Treatment



Analysis of deviations from type. Our expectation is that subjects will perform quite consistently at the RET over multiple rounds of the experiment. In the lab experiments rarely did each subject's performance deviate significantly from their performance "type" (either low performance types deviating significantly above the mean or high performance types deviating significantly below the mean). As an illustration, Duch and Solaz (2016) defined low performance deviations as an outcome in which a low performance type performed more than 12 additions. And high performance deviations occurred when a high performance type only manages less than 10 correct additions. In the lab results there were only 78 occurrences of the high performance deviations which represents less than 5 percent of high performer outcomes. And there are only 68 occurrences of low performance deviations which also represents less than 5 percent of low performer outcomes. We will replicate the same analysis with the online synchronic experiment.

Our contention is that one's performance type explains cheating behaviour. Hence cheating behaviour should be consistent within subjects – it should not fluctuate significantly and we would not expect it to respond to stochastic shocks in performance. This implies that, for any particular subject, cheating behaviour is not correlated with RET outcomes that deviate significantly from their overall performance levels. Its not the case, for example, that when a high performance type experiences an unexpectedly poor RET performance her cheating sharply declines. To test this, Duch and Solaz (2016) calculated, for occurrences of both low and high performance deviations, the average change in percent of earnings evaded. In both cases, the average change was not significantly different from zero.[2] Again a similar analysis will be conducted with the online experiment data.

[2] Figure 8 in the Appendix provides a frequency plot of these deviations in cheating for the low and high performance deviation cases.

Risk Preferences The performance type effect estimated by Duch and Solaz (2016) was not confounded with risk preferences. The last module of the lab experiment consisted of a lottery-choice test consisting of ten pairs, which is based in the low-payoff treatment studied in Holt and Laury (2002). The lottery choices are shown in Table 5 in the Appendix. Subjects indicated their preferences, choosing Option A or Option B, for each of the ten paired lottery choices, and they know one of these choices would be selected at random ex post and played to determine the earnings for the option selected. The Duch and Solaz (2016) measure of risk aversion was simply the sum of the safe choices by each subject.

The overall correlation between performance (number of correct additions) and the sum of safe lottery choices in the lab experiment was -0.03. Hence Duch and Solaz (2016) found little evidence that high performance types exhibit high degrees of risk aversion (or vice versa for that matter). Duch and Solaz (2016) also estimated the treatment effects controlling for low, medium, and high levels of risk aversion. A version of Figure 3, controlling for risk aversion, is presented in Figure 10 in the Appendix. Within all of the risk aversion categories, high performance types consistently cheat more than low performance types; and for the most part the differences in means are statistically significant.[3] We would conduct similar analysis of the online synchronic experimental data. Note that we include the identical risk preference measure in the online experiment that was administered in the lab experiment.

3.3 Multivariate

Cheating will be measured in two fashions in the analysis of the experimental results: 1) simply whether or not the subject correctly reported her income from the real effort

[3]There is one exception in which low and high performance types have essentially the same average rates of cheating. The cells sizes get relatively small which accounts for some of the imprecise estimates. In addition there are empty cells – these are typically for the risk seeking category.

task; and 2) the percent of a subject's actual income from the real effort task that was not reported. We will model whether or not subjects cheated as a dichotomous variable. Subjects who reported their actual winnings will be coded zero and those who reported amounts that deviated from their actual winnings are coded one. Results for the dichotomous measure of cheating from the Duch and Solaz (2016) lab experiment are reported in Table 2.

Model 1 in Table 2 presents the Baseline treatment results from the Duch and Solaz (2016) experiment in which all subjects earn the same amount for each correct addition (10 pence). This Baseline treatment represents a context in which the labour market rewards ability and there are no structural factors that cause some individuals to earn more for their (equal) ability than others. Model 1 includes a measure of Performance (Correct Additions) and the cost of full tax compliance. For this Baseline treatment our principal expectation is that performance should be correlated with cheating – better performers (more correct additions) should cheat more. This is precisely what we see in the Duch and Solaz (2016) lab results. The positive statistically significant coefficient on Performance indicates that the probability of cheating rises with performance. And also as expected the coefficient on the cost of tax compliance is positive and significant. Hence in contexts where ability is the sole factor determining income, high ability types are more likely to cheat than those with lower abilities or effort. And this holds when Duch and Solaz (2016) control for the cost of tax compliance which, as one would expect, is also positively correlated with cheating.[4] This Baseline multivariate analysis will be replicate for the data from the online synchronic experiment.

[4]Table 3 in Appendix 1 replicates the analysis in Table 2 with a dependent variable that measures the percent of a subject's earnings that were not reported. The results essentially confirm the findings from Table 2.

Table 2: Probit model of cheating regressed on performance

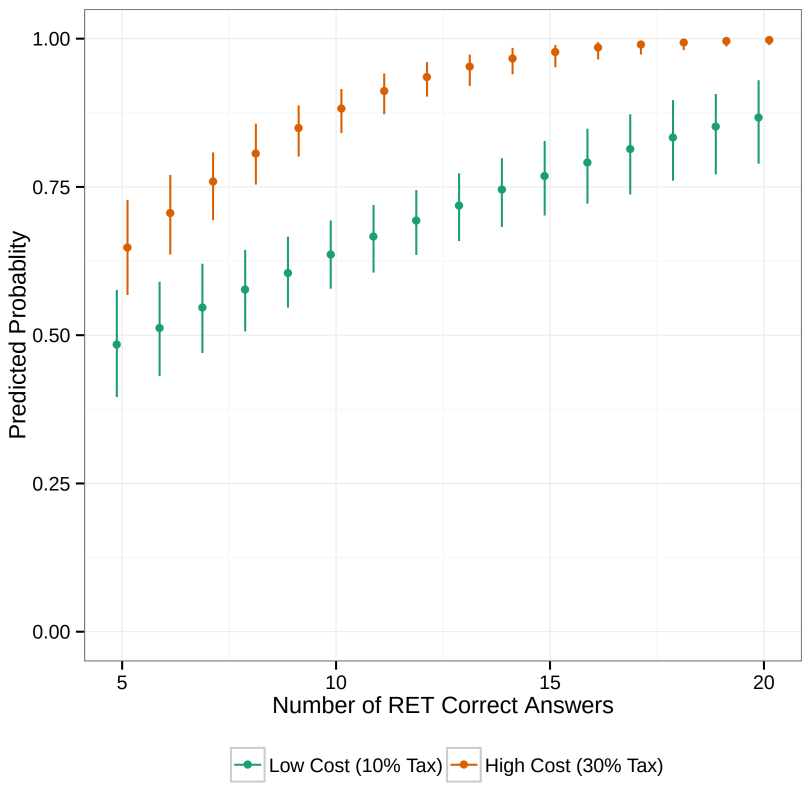
	Baseline
# of Additions	0.036^{**}
	(0.018)
Cost of Compliance	0.003^{***}
	(0.0004)
Constant	-0.428^{**}
	(0.186)
Observations	720
Log Likelihood	-319.530
Akaike Inf. Crit.	645.061

Note: *p<0.1; **p<0.05; ***p<0.01

Predicted effects. Figure 4 helps understand the estimated effects by presenting predicted cheating probabilities from the results in Table 2. For the Baseline treatment, the graphs plot the predicted probability of cheating against subjects' performance in the real effort task. The graphs on the left are for the low-cost condition (a tax rate of 10 percent) and on the right for the high cost condition (a tax rate of 20 percent). In all treatments there are statistically significant positive relationships between performance and cheating.

The Baseline results are perfectly consistent with our two primary conjectures. First, there is a strong positive relationship between Performance and cheating. And, consistent with our second conjecture, cheating is higher in the high cost condition. Nevertheless, even in the high cost condition, this positive relationship between Performance and cheating persists.

Figure 4: Predicted Probabilities of Cheating for Treatments: 10% and 30% Tax Rates

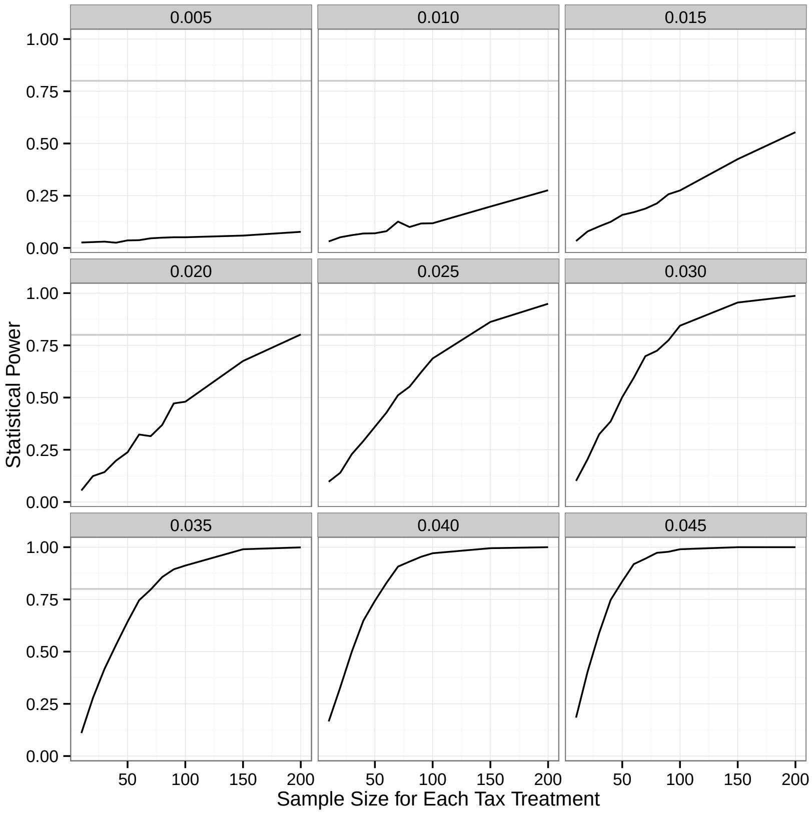


4 Power Analysis

Based on the results of Table 2 and average rate of compliance in the data, we calculate the statistical power of our experimental design. In order to calculate the statistical power of such situation, we rely on Monte Carlo methods. We randomly sample the RET performance variable from the empirical distribution of RET performance of the Lab subjects assigned to the Baseline treatment. We have other hypothetical coefficients fixed at the estimated level: the number of addition at 0.036 and the constant term at -0.428. The variable of interest, the hypothetical coefficient size for the number of correct answers are varied from 0.005 to 0.045 incremented by 0.005. We also assume that there is an equal number of subjects for each of two tax-rates, 10% and 30%. The number of experimental subjects in each group is varied from 10 to 200. As each subject is going to play ten rounds of the tax compliance game, the total number in each simulation is (Number of subjects) $\times$ (Number of rounds) $\times$ (Number of different tax rates). Figure 5 shows the result of power calculation. For each setting, we have run 1000 simulations, estimated a probit model, and checked the proportion of coefficients where the coefficient for the cost of compliance is significant. The horizontal axis is sample size, and The confidence level is 95% (two-tail). Each panel corresponds to the hypothetical effect size. If the effect size is close to the lab results, which is 0.036, the sample size of our study has enough statistical power. The minimum effect size we can detect with the statistical power of 0.8 is 0.03.

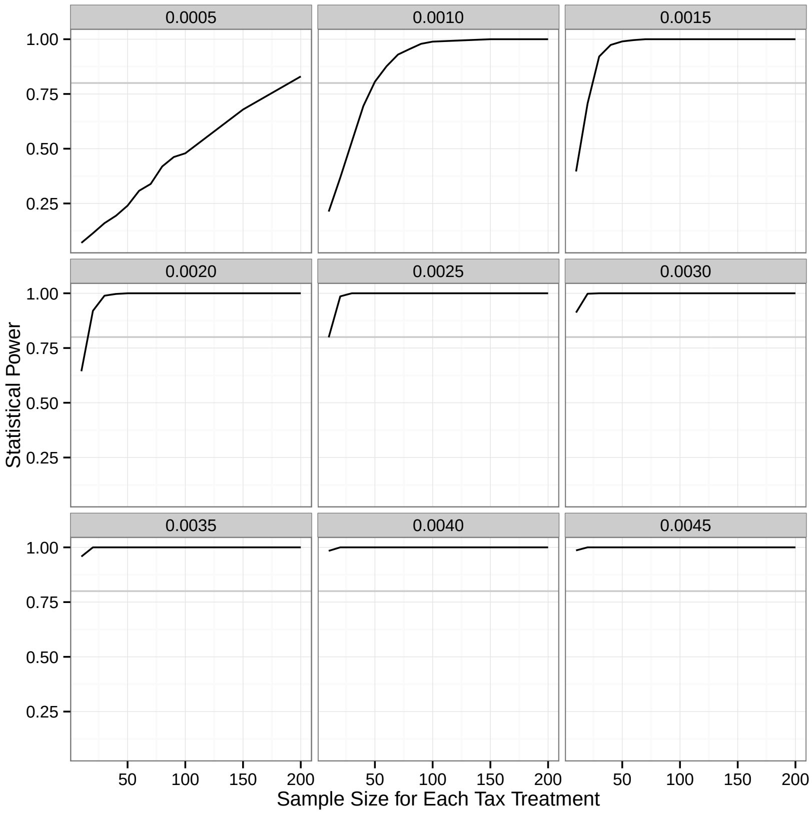
Another effect of interest is the cost of compliance: the variable is calculated as the multiplication of tax rate and RET performance. In Figure 6, we conduct the power analysis for the effect of number of correct answers. This time, we fix the coefficient for the number of correct answers at the estimated level of 0.036, and then varied the hypothetical effect size of cost of compliance from 0.0005 to 0.0045. Other settings are the same as the first power analysis.

Figure 5: Power calculations for difference hypothetical effect size of number of correct answers



As the figure indicates, the statistical power of our analysis is every strong when the number of subjects in each tax rate treatment is 100. The minimum effect we can detect with the sample size is 0.001 which is a conservative assumption considering the fact that this is a third of the size we have found in Duch and Solaz (2016). If the true effect size is smaller than 0.001, there is a chance that we are able to detect the effect. The statistical power at 100 subjects is around 0.5, so we will not be able to detect the effect in a half of case. In our other paper on comparing student subjects in CESS lab and MTurk workers, we have found that MTurk workers exhibit similar effects of treatment effects than lab sujbest but with a smaller size (Beramendi, Duch and Matsuo 2016). This would necessitate that we would need more sample size larger than the lab. The number of lab subjects in each tax rate treatment for this baseline condition is 24, and

Figure 6: Power calculations for difference hypothetical effect size of cost of compliance



having 100 for online settings will be enough to detect the treatment effects.

References

Balietti, Stefano. 2015. "nodeGame: Real-Time Social Experiments in the Browser." Working Paper.

Beramendi, Pablo, Raymond M. Duch and Akitaka Matsuo. 2016. "When Lab Subjects Meet Real People: Comparing Different Modes of Experiments.".

Duch, Raymond and Hector Solaz. 2016. "Why we Cheat: Experimental Evidence on Tax Compliance." Working Paper. Centre for Experimental Social Sciences, Nuffield College, University of Oxford.

Holt, Charles A. and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92:1644–1655.

Levitt, Steven D. and John A. List. 2007. "What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21:153–74.

APPENDIX

Appendix 1: Regression on Percent of Earnings not Reported

Table 3 replicates the analysis in Table 2 with a dependent variable that measures the percent of a subject's earnings that were not reported. The results essentially confirm the findings from Table 2. In the baseline treatment in which monetary rewards are strictly determined by performance and ability, our two initial conjectures are supported: there is a significant positive correlation between earnings and cheating; and cheating rises significantly with the cost of compliance, i.e., earnings multiplied by the tax rate.

Table 3: Percent Evaded Regressed on Performance

	Baseline
# of Additions	0.025
	(0.016)
Cost of Compliance	0.002***
	(0.0003)
ϕ	0.409
	(0.017)
Observations	720
Log Likelihood	−319.530
Akaike Inf. Crit.	645.061

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix 2: Probit regression using first round of each session

A conservative test of our argument regarding ability and cheating is to estimate the probit regression models only employing observations from the first round of each of the tax compliance experimental sessions. Table 4 presents the results. This essentially replicates Table 2 from the main text – the only difference is that the estimates are based only on the decisions taken by the subjects in the first round of each session. Having fewer observations results in less precise estimates but also reduces the amount of information available to distinguish between the highly correlated Addition and Cost of Compliance variables.

Nevertheless, the results in Table 4 essentially confirm our argument that ability plays an important role in explaining cheating behaviour. In most of the models, performance is positively, and significantly, correlated with cheating. The one, important, exception is in the Baseline model in which the Addition variable is insignificant. When we estimate the Baseline model without the Cost of Compliance control the coefficient on Addition is positive and weakly significant at the 0.1 level. On balance though its clear that, controlling for the Cost of Compliance, ability is positively correlated with cheating.

It might be the case that high ability types only begin to cheat aggressively because they observe that they are net contributors to the public good. However, it does not seem to be the case that this correlation is an artefact of learning over the course of the 10 sessions each subject plays. The correlation between ability and cheating seems to emerge at the outset of the game.

Table 4: Regression using first round of each session

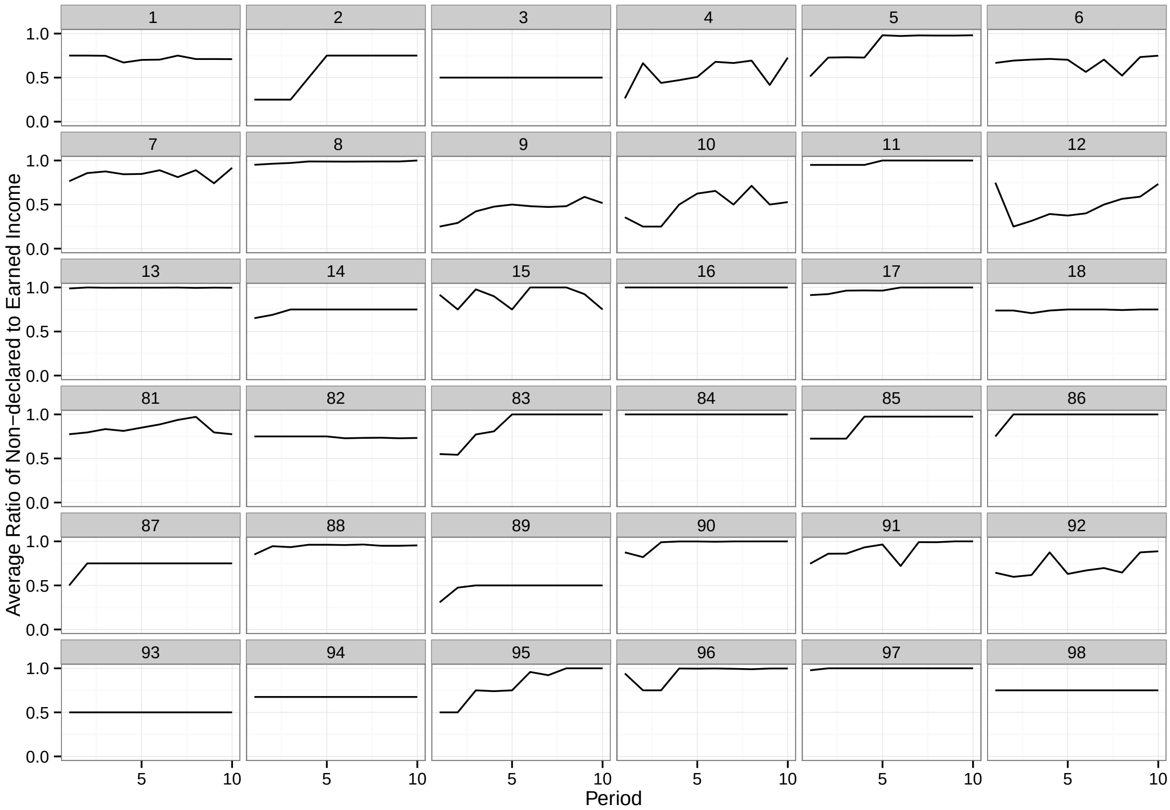
	Baseline
# of Additions	0.004
	(0.053)
Cost of Compliance	0.004^{***}
	(0.002)
Constant	-0.473
	(0.535)
Observations	72
Log Likelihood	-33.437
Akaike Inf. Crit.	72.874

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix 3: Average Cheating Rates over Each Round within Session

Subjects within a particular group earn and report income over ten periods. For each of these groups, Figure 7 presents the average cheating rates for each of the ten periods. Average cheating in each group is for the most part quite constant. And to the extent that there is a trend in cheating, it is consistent with trends we typically find in public goods games whereby contributions fall as the subjects approach the final period of play (Levitt and List 2007). Of the 36 groups, there are around ten groups for which average cheating rates rise significantly by period ten. Hence there is some evidence here to suggest that any obligation subjects feel toward compliance erodes significantly as the end of the game approaches – in a number of cases reaching levels predicted by equilibrium reasoning.

Figure 7: Ratio of Non-declared to Total Earnings Averaged over Subjects

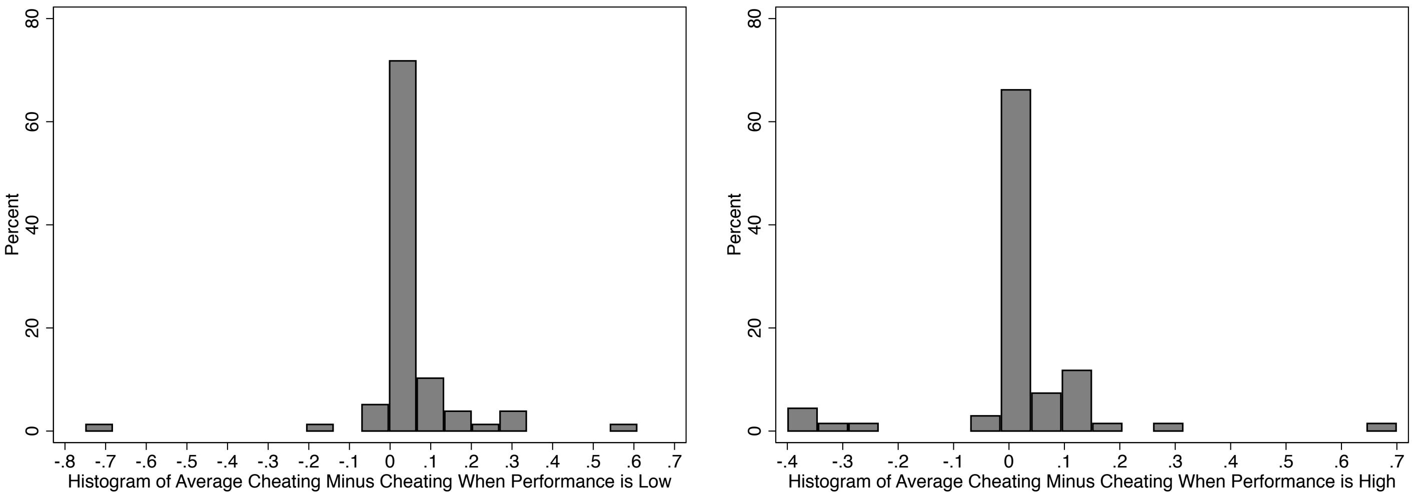


Appendix 4: Cheating Behaviour for Low and High Performance Types When Performance Deviates Significantly from their Average Performance

Our contention is that the RETs identify performance types in the population – individuals who recognise themselves as having ability versus not having high ability levels. The expectation we note in the text is that performance is likely to be very consistent within subjects. As an illustration, we define low performance deviations as an outcome in which a low performance type performs more than 12 additions. And high performance deviations occur when a high performance type only manages less than 10 correct additions. There are only 78 occurrences of the high performance deviations which represents less than 5 percent of high performer outcomes. And there are only 68 occurrences of low performance deviations which also represents less than 5 percent of low performer outcomes.

Our contention is that one's performance type explains cheating behaviour. Hence cheating behaviour should be very consistent within subjects – it should not fluctuate significantly and we would not expect it to respond to stochastic shocks in performance. This implies that, for any particular subject, cheating behaviour is not correlated with RET outcomes that deviate significantly from their overall performance levels. Its not the case, for example, that when a high performance type experiences a negative shock her cheating sharply declines. For both the low and high performance deviations we calculated the average change in percent of earnings evaded. Figure 8 provides a frequency plot of these deviations in cheating for the low and high performance deviation cases. As is clear in Figure 8, for both cases, the average change was not significantly different from zero.

Figure 8: Deviations in Performance and Cheating for Low and High Performance Types



(a) High Performance Types

(b) Low Performance Types

Appendix 5: Risk Aversion

The fourth and last module of the experiment consists of a lottery-choice test consisting of ten pairs, which is based in the low-payoff treatment studied in (Holt and Laury 2002). The lottery choices (shown in Table 5) are structured so that the crossover point to the high-risk lottery can be used to infer the degree of risk aversion. Subjects indicate their preferences, choosing Option A or Option B, for each of the ten paired lottery choices, and they know one of these choices would be selected at random ex post and played to determine the earnings for the option selected.

Table 5: Lottery Choices (in the Lab)

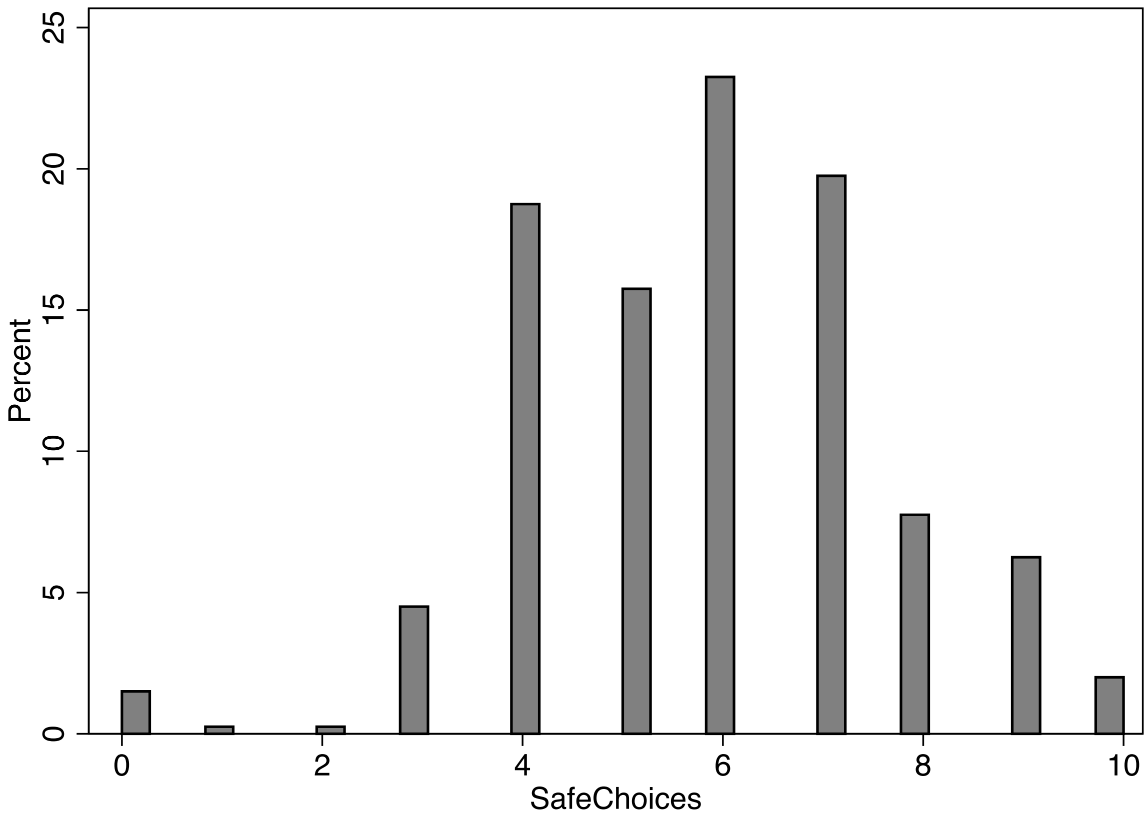
	Option A	Option B
1	10% of 2.00£, 90% of 1.60£	10% of 3.85£, 90% of 0.10£
2	20% of 2.00£, 80% of 1.60£	20% of 3.85£, 80% of 0.10£
3	30% of 2.00£, 70% of 1.60£	30% of 3.85£, 70% of 0.10£
4	40% of 2.00£, 60% of 1.60£	40% of 3.85£, 60% of 0.10£
5	50% of 2.00£, 50% of 1.60£	50% of 3.85£, 50% of 0.10£
6	60% of 2.00£, 40% of 1.60£	60% of 3.85£, 40% of 0.10£
7	70% of 2.00£, 30% of 1.60£	70% of 3.85£, 30% of 0.10£
8	80% of 2.00£, 20% of 1.60£	80% of 3.85£, 20% of 0.10£
9	90% of 2.00£, 10% of 1.60£	90% of 3.85£, 10% of 0.10£
10	100% of 2.00£, 0% of 1.60£	100% of 3.85£, 0% of 0.10£

Table 6: Lottery Choices (in the Lab)

	Option A	Option B
1	10% of 1.00$, 90% of 0.80$	10% of 1.92$, 90% of 0.05$
2	20% of 1.00$, 80% of 0.80$	20% of 1.92$, 80% of 0.05$
3	30% of 1.00$, 70% of 0.80$	30% of 1.92$, 70% of 0.05$
4	40% of 1.00$, 60% of 0.80$	40% of 1.92$, 60% of 0.05$
5	50% of 1.00$, 50% of 0.80$	50% of 1.92$, 50% of 0.05$
6	60% of 1.00$, 40% of 0.80$	60% of 1.92$, 40% of 0.05$
7	70% of 1.00$, 30% of 0.80$	70% of 1.92$, 30% of 0.05$
8	80% of 1.00$, 20% of 0.80$	80% of 1.92$, 20% of 0.05$
9	90% of 1.00$, 10% of 0.80$	90% of 1.92$, 10% of 0.05$
10	100% of 1.00$, 0% of 0.80$	100% of 1.92$, 0% of 0.05$

Results for the safe choice measure are presented in Figure 9. This represents the distribution of subjects from all 14 sessions we report on here. We create risk preference categories in order to assess whether controlling for risk preferences significantly affects the high versus low performance effects we present in the text. Subjects are categorised into three risk categories based on the sum of the safe choices they make: risk seeking (a sum less than 4); risk neutral (a sum between 4 and 6); risk averse (a sum greater than 6).

Figure 9: Distribution of Sum of Safe Choices in Lottery Game



In the text, Figure 3 presents the difference in reported earnings for high versus low performance types – these differences are presented for each of the different treatment sessions. Again, for each of the treatment sessions, Figure **??** compares the high versus low performance type differences within each of the three risk categories. Results for the baseline treatment session indicate that the relationship between performance type and cheating for the most part persists when we control for risk preferences. For risk seeking and risk neutral subjects there is a strong positive relationship between performance type and cheating. Amongst the most risk averse subjects, though, cheating rates are similar for both high and low performance types. For the most risk averse subjects, regardless of performance type, it would seem that uncertainty regarding the choices of other subjects results in very high levels of cheating.

For subjects in the "status" treatment sessions, we also find that our estimated relationship between performance type and cheating is robust to the introduction of controls for risk preferences. This holds for risk neutral and risk averse subjects in both the low and high "status" treatments. In this case there are insufficient observations in the risk seeking categories to make any comparisons. We see a similar pattern for subjects in the "shock" treatment sessions – there is a positive relationship between performance type and cheating for risk neutral and risk averse subjects in both the control and "shock" treatments. Finally Figure 10 presents the results when we introduce the risk preference controls for those in the "redistribute" treatment sessions. Here we find a weaker, although still positive, relationship between performance and cheating for risk neutral and risk averse subjects.

Figure 10: Percent Evaded: High and Low Performance Types and Risk

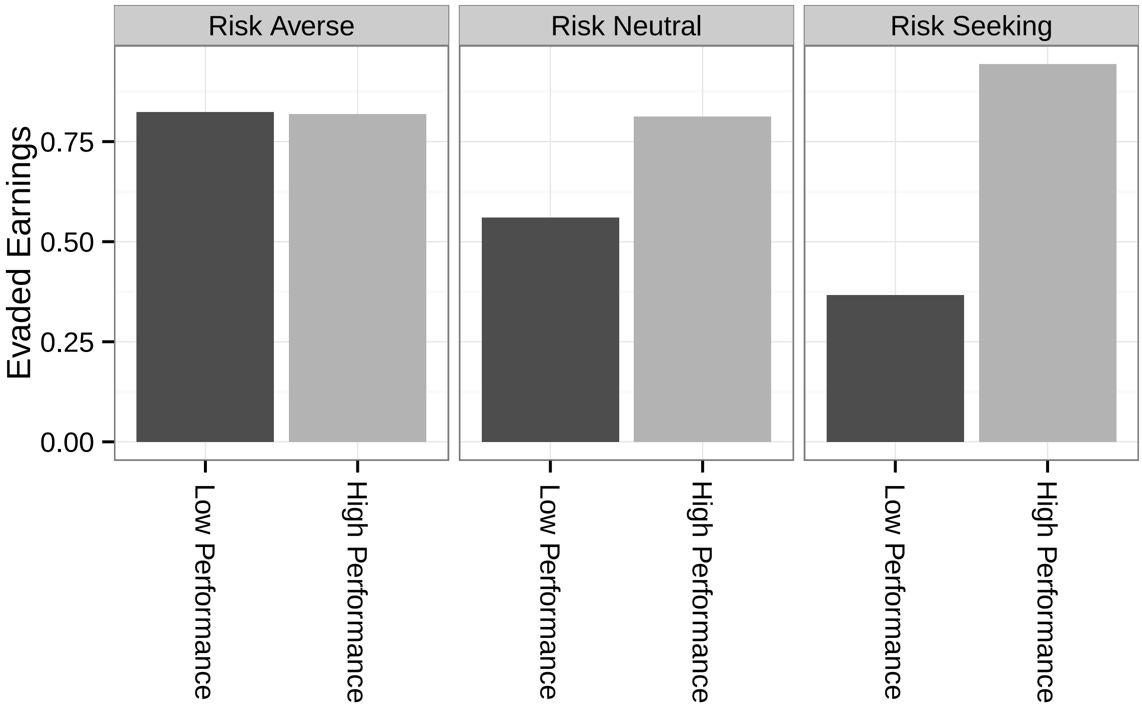


Table 2 in the main text presented the multivariate probit regression results for each of the treatment session models. Table 7 re-estimates these models and includes in the estimation the safe choice risk averse variable described in Figure 9.

Table 7: Probit model of cheating regressed on performance (with safe choice)

	Baseline
# of Additions	0.036^{**}
	(0.018)
Cost of Compliance	0.003^{***}
	(0.0004)
Constant	-0.586^{**}
	(0.262)
Observations	720
Log Likelihood	-319.151
Akaike Inf. Crit.	646.302

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix 6: Instructions for Treatment Sessions

Figure 11: Instructions for Baseline Treatment

NUFFIELD CENTER FOR EXPERIMENTAL SOCIAL SCIENCES

Module 2:

a) This module consists of 10 rounds. At the beginning of the module participants are randomly assigned to groups of four. You won't know the identity of the other group members. The composition of each group will remain unchanged.

b) In each one-minute round you will be asked to compute a series of additions. Your *Preliminary Gains* depend on how many correct answers you provide. You will get **150 ECUs** for each correct answer.

c) At the end of each round, once you have received information concerning your *Preliminary Gains*, you will be asked to declare these gains. In this module **10%** of these *Declared Gains* will then be deducted from your *Preliminary Gains.*

d) In each round there is a certain probability that your *Declared Gains* will be compared with your actual *Preliminary Gains* in order to verify these two amounts correspond. In this module this probability is **0%.**

e) If this verification finds a discrepancy between the *Preliminary* and *Declared* gains an extra amount will be deducted from your *Preliminary Gains*. In this module this amount will correspond to **50%** of the observed discrepancy. In addition, the regular deduction of **10%** will apply to the *Preliminary Gains* and not to the declared amount.

f) Deductions applying to the four group members will then be pooled and equally distributed amongst those members.

g) Your profits are calculated and displayed at the end of each round in the following manner:

Profit = Preliminary Gains – Deduction from the Declared Gains – Potential deductions due to discrepancy + Group amount

h) At the end of the module one round will be chosen at random, and your earnings will be based on your profit of that round at the exchange rate **300ECUs = 1 £**

i) You will be informed of your earnings for this module at the end of the experiment.

Figure 12: Instructions for Status Treatment

NUFFIELD CENTER FOR EXPERIMENTAL SOCIAL SCIENCES

Module 2:

a) This module consists of 10 rounds. At the beginning of the module participants are randomly assigned to groups of four. You won't know the identity of the other group members. The composition of each group will remain unchanged. In each group two members will be Type G and the other two will be Type K. Every participant has an equal chance to be designated as either type, G or K.

b) In each one-minute round you will be asked to compute a series of additions. Your *Preliminary Gains* depend on how many correct answers you provide. If you are Type G you will get **200 ECUs** for each correct answer, while if you are Type K you'll get **100 ECUs**.

c) At the end of each round, once you have received information concerning your *Preliminary Gains*, you will be asked to declare these gains. In this module **10%** of these *Declared Gains* will then be deducted from your *Preliminary Gains.*

d) In each round there is a certain probability that your *Declared Gains* will be compared with your actual *Preliminary Gains* in order to verify these two amounts correspond. In this module this probability is **0%.**

e) If this verification finds a discrepancy between the *Preliminary* and *Declared* gains an extra amount will be deducted from your *Preliminary Gains*. In this module this amount will correspond to **50%** of the observed discrepancy. In addition, the regular deduction of **10%** will apply to the *Preliminary Gains* and not to the declared amount.

f) Deductions applying to the four group members will then be pooled and equally distributed amongst those members.

g) Your profits are calculated and displayed at the end of each round in the following manner:

Profit = Preliminary Gains – Deduction from the Declared Gains – Potential deductions due to discrepancy + Group amount

h) At the end of the module one round will be chosen at random, and your earnings will be based on your profit of that round at the exchange rate **300ECUs = 1 £**

i) You will be informed of your earnings for this module at the end of the experiment.

Figure 13: Instructions for Redistribute Treatment

NUFFIELD CENTER FOR EXPERIMENTAL SOCIAL SCIENCES

Module 2:

a) This module consists of 10 rounds. At the beginning of the module participants are randomly assigned to groups of four. You won't know the identity of the other group members. The composition of each group will remain unchanged.

b) In each one-minute round you will be asked to compute a series of additions. Your *Preliminary Gains* depend on how many correct answers you provide. You will get **150 ECUs** for each correct answer.

c) At the end of each round, once you have received information concerning your *Preliminary Gains*, you will be asked to declare these gains. In this module **10%** of these *Declared Gains* will then be deducted from your *Preliminary Gains.*

d) In each round there is a certain probability that your *Declared Gains* will be compared with your actual *Preliminary Gains* in order to verify these two amounts correspond. In this module this probability is **0%.**

e) If this verification finds a discrepancy between the *Preliminary* and *Declared* gains an extra amount will be deducted from your *Preliminary Gains*. In this module this amount will correspond to **50%** of the observed discrepancy. In addition, the regular deduction of **10%** will apply to the *Preliminary Gains* and not to the declared amount.

f) Deductions applying to the four group members will then be pooled. If you are one of the two members of the group who has solved the most additions, you will be considered Type K and will receive 15% of the pooled deductions. If not (you are one of the two members of the group who has solved the least additions), you will be considered Type G and will receive 35% of the pooled deductions. When group members are tied in the number of additions solved, assignment to Type G and K will be decided randomly. Assignment to Types G and K will be determined each round.

Figure 14: Instructions for Shock Treatment

NUFFIELD CENTER FOR EXPERIMENTAL SOCIAL SCIENCES

Module 2:

a) This module consists of 10 rounds. At the beginning of the module participants are randomly assigned to groups of four. You won't know the identity of the other group members. The composition of each group will remain unchanged.

b) In each one-minute round you will be asked to compute a series of additions. Your *Preliminary Gains* depend on how many correct answers you provide. You will get **100 ECUs** for each correct answer.

c) In each group, each round, two members will receive a *Bonus* of **1300 ECUs**. Every participant has an equal chance to receive the *Bonus* each round. This Bonus is added to your *Preliminary Gains* in case you receive it.

d) At the end of each round, once you have received information concerning your *Preliminary Gains*, you will be asked to declare these gains. In this module **10%** of these *Declared Gains* will then be deducted from your *Preliminary Gains.*

e) In each round there is a certain probability that your *Declared Gains* will be compared with your actual *Preliminary Gains* in order to verify these two amounts correspond. In this module this probability is **0%.**

f) If this verification finds a discrepancy between the *Preliminary* and *Declared* gains an extra amount will be deducted from your *Preliminary Gains*. In this module this amount will correspond to **50%** of the observed discrepancy. In addition, the regular deduction of **10%** will apply to the *Preliminary Gains* and not to the declared amount.

g) Deductions applying to the four group members will then be pooled and equally distributed amongst those members.

h) Your profits are calculated and displayed at the end of each round in the following manner:

Profit = Preliminary Gains – Deduction from the Declared Gains – Potential deductions due to discrepancy + Group amount

i) At the end of the module one round will be chosen at random, and your earnings will be based on your profit of that round at the exchange rate **300ECUs = 1 £**

j) You will be informed of your earnings for this module at the end of the experiment.

Appendix 6: Screen Shots from Real Effort Tasks

4.0.1 Lab

Figure 15: Screen Shot One from Real Effort Task

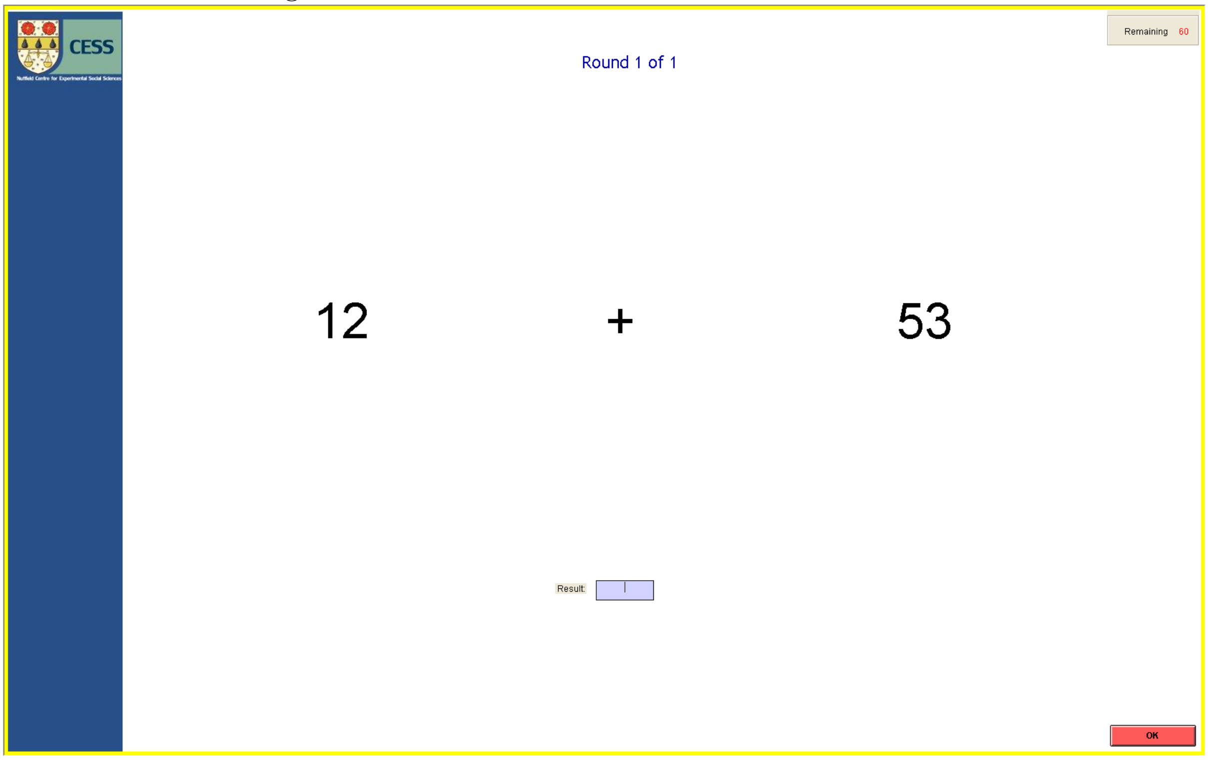


Figure 16: Screen Shot Two from Real Effort Task

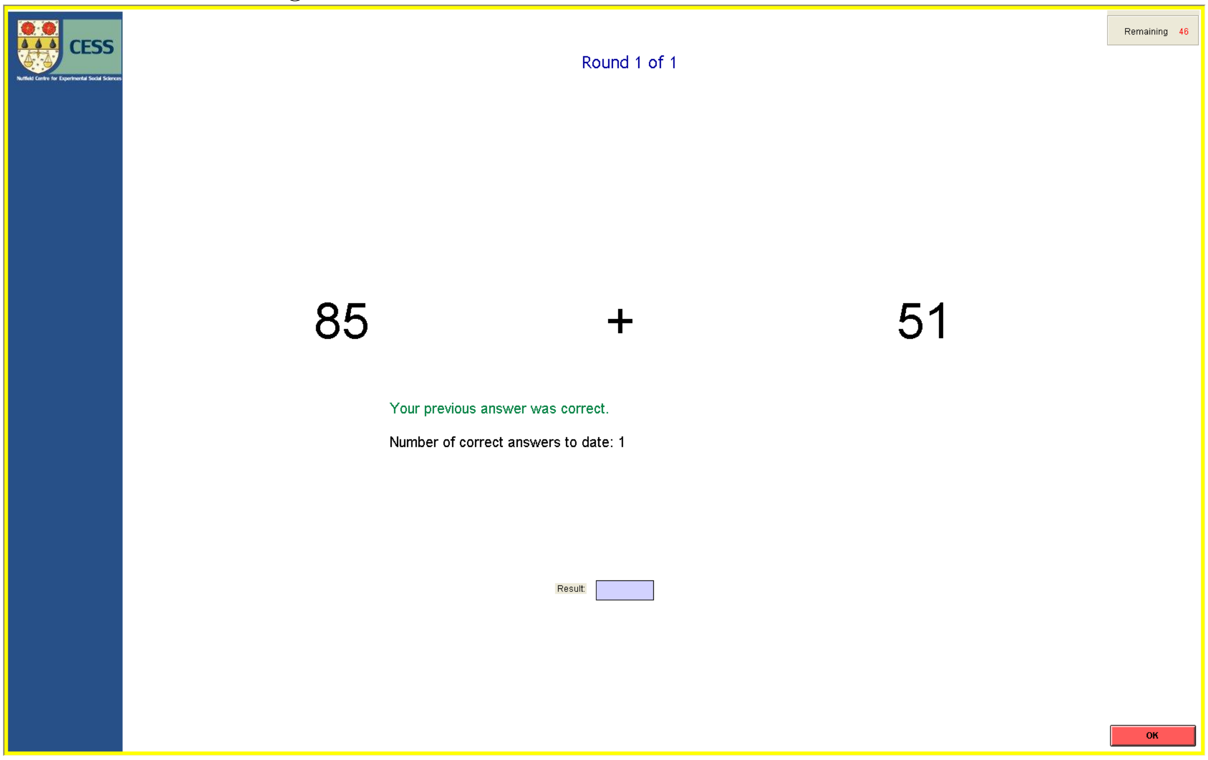


Figure 17: Screen Shot Three from Real Effort Task

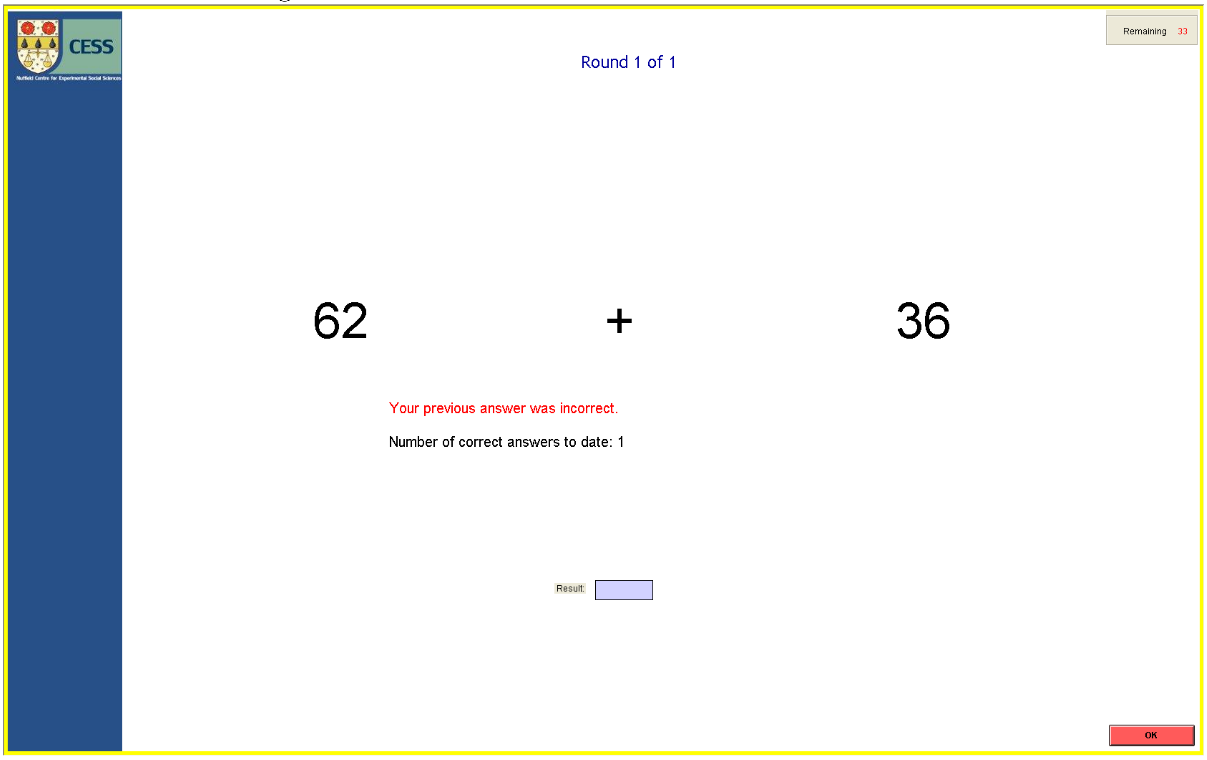


Figure 18: Screen Shot Four from Real Effort Task

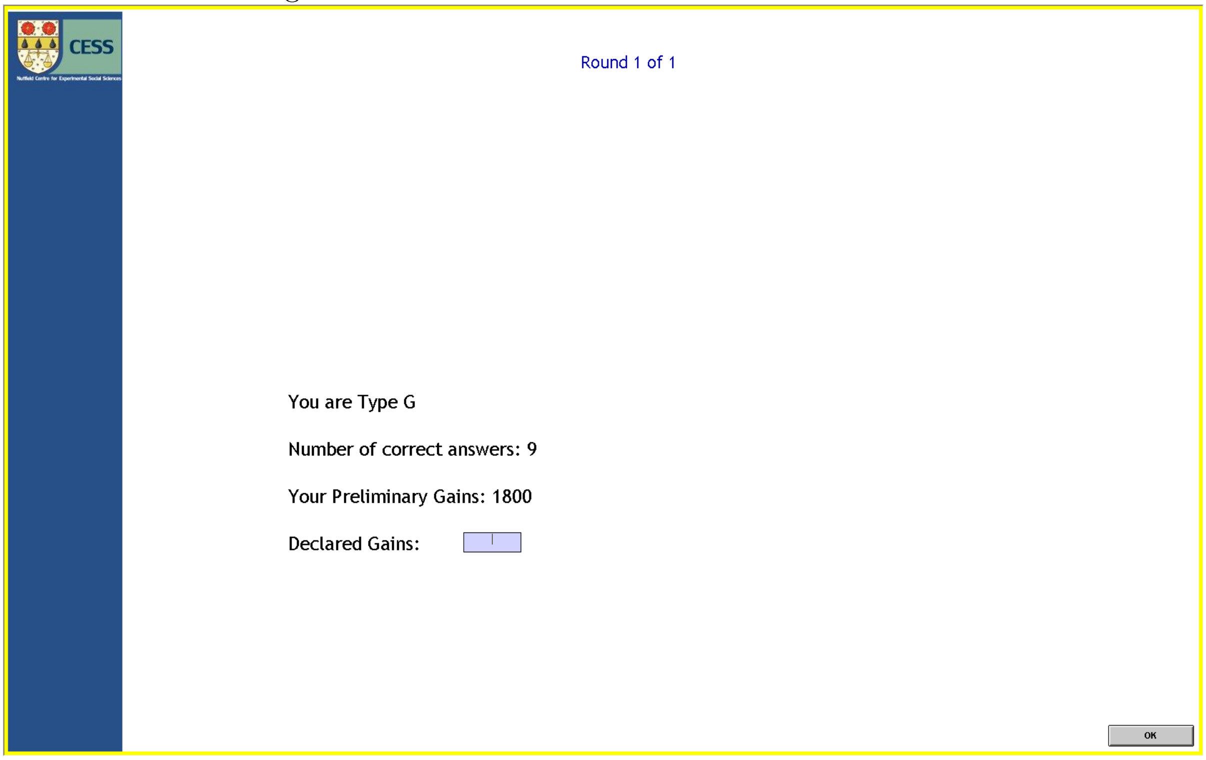
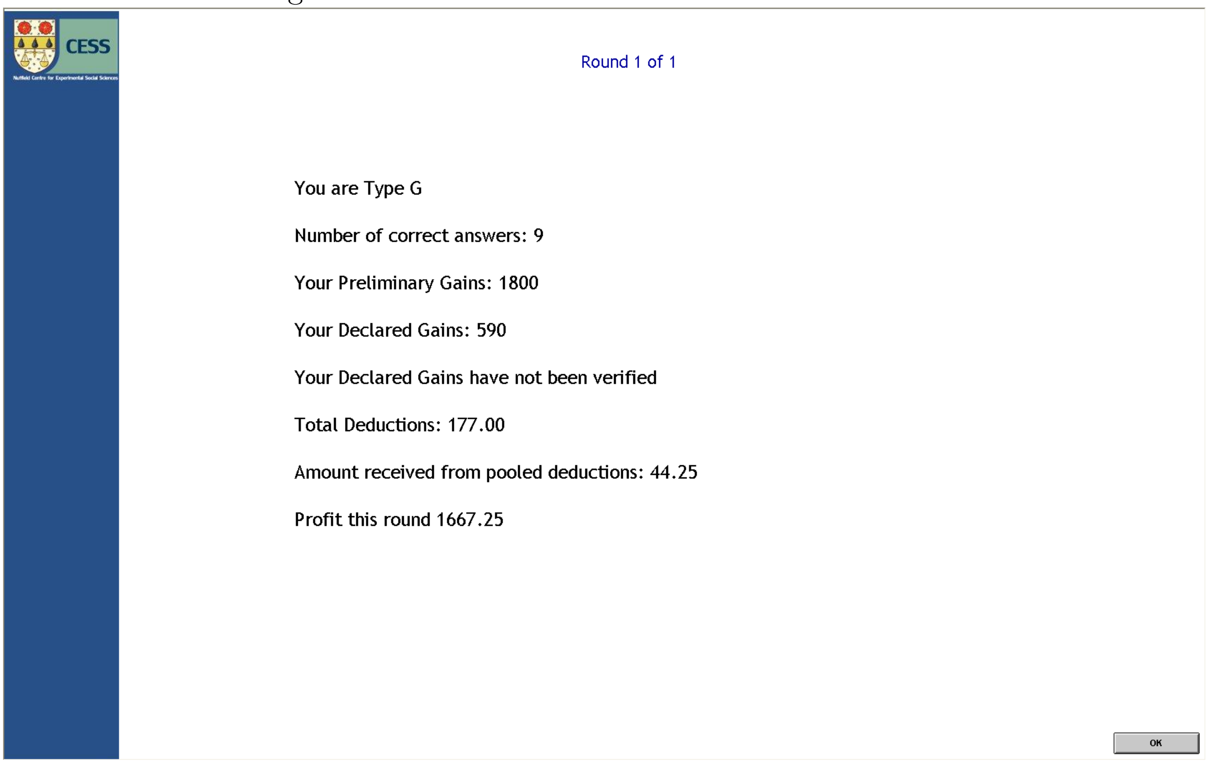


Figure 19: Screen Shot Five from Real Effort Task



4.0.2 Online

Figure 20: Screen Shot One from Real Effort Task

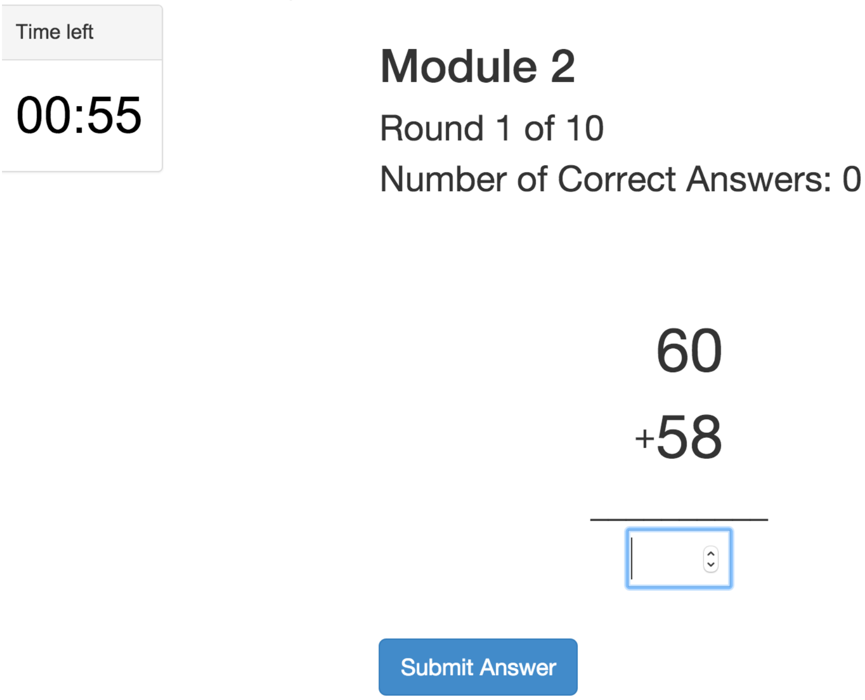


Figure 21: Screen Shot Two from Real Effort Task

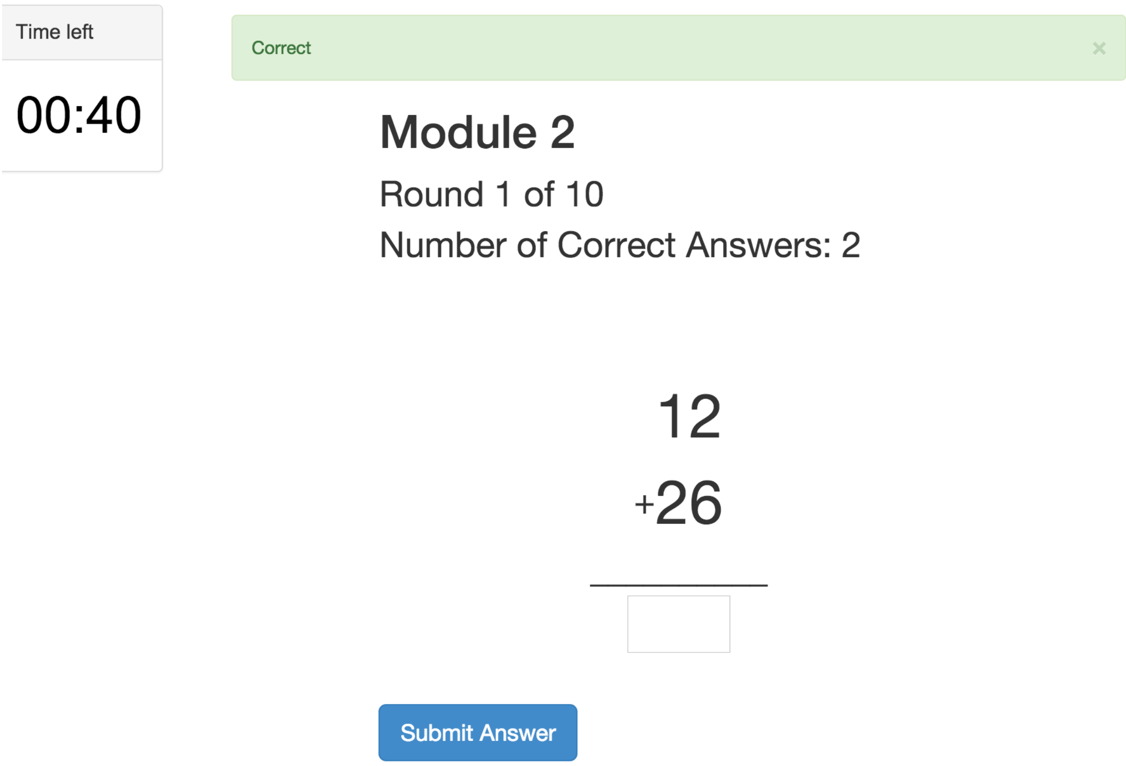


Figure 22: Screen Shot Three from Real Effort Task

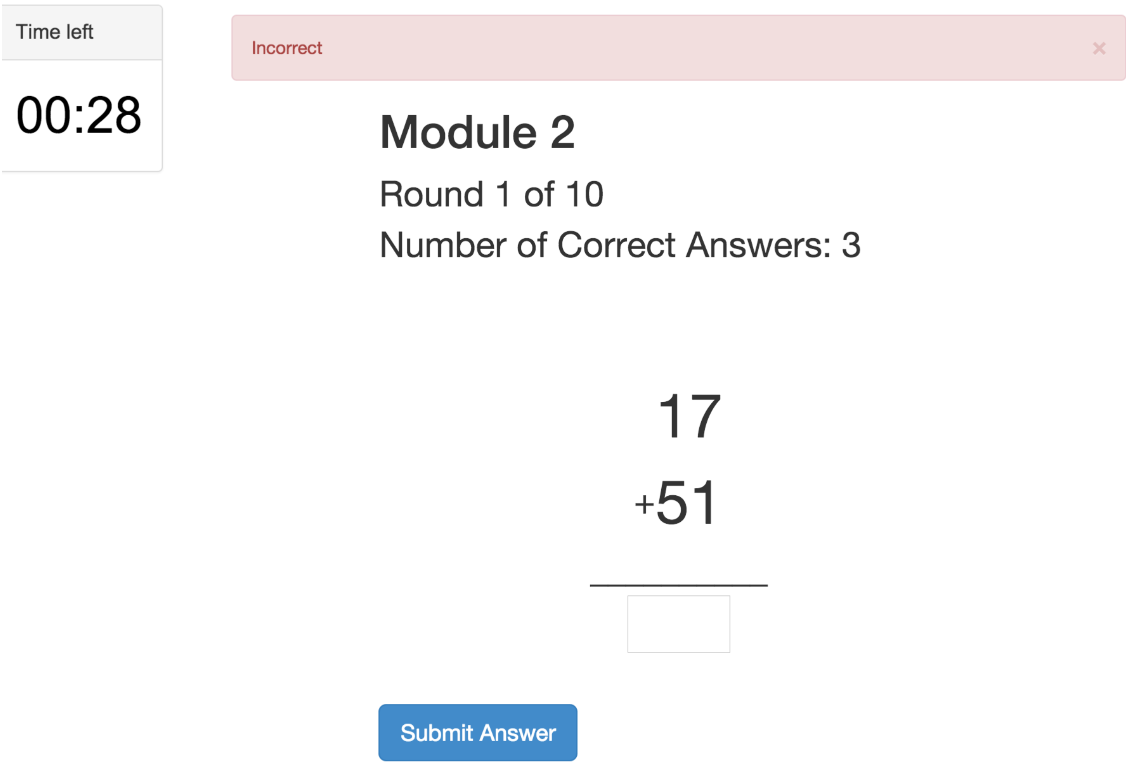


Figure 23: Screen Shot Four from Real Effort Task

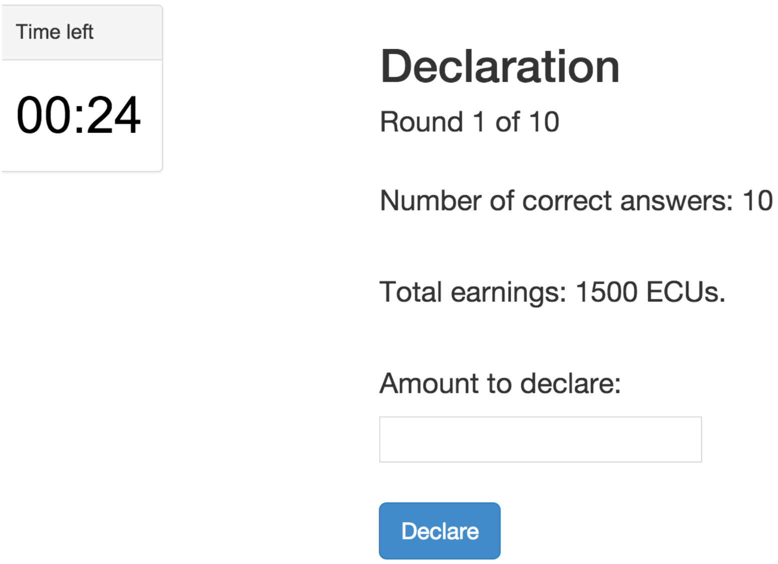


Figure 24: Screen Shot Five from Real Effort Task